Architecture of Estonia

Elements of a digital society

Andres Kütt

October 21, 2017

Cybernetica, architect

Introduction

Andres Kütt

- Building software for money since 1993
- An architect of some capacity for the past 15 years
- ≈MSc (UT, Statistics), MBA (EBS), MSc (MIT)
- Currently architect at Cybernetica
- Done Skype, a few banks, Estonian government etc.
- Variety of courses and seminars in various schools in Estonia and abroad

Housekeeping

- Three larger blocks of teaching
  - Divided into three 20-25 minute chunks each
  - Punctuated by joint discussion
- Permitted are
  - Questions including questioning whatever I tell you
  - Moderate sharing of personal experience
  - Arrival and leaving whenever
- Wasting time is not OK. Especially your own

Main goals of today

We are limited by the time we have, the following should be possible to accomplish:

- Provide a minimal set of theory
- Pick two important concepts and discuss them in some detail relying on theory
- Show how the theory and the concepts related to success and failure in Estonia
- Answer any questions you might have about Estonian digital infrastructure

The goal is to provide a foundation for you to rapidly build on

Not too specific and not too theoretical.

I'll do my best to find the balance

Structure of today

- General theory
  - Introduction to the day, the lecturer. Main goals
  - Key mental models for thinking of digital government
  - Ecosystems. Their importance and basic principles
- Fundamental concepts
  - Privacy. Information, data and interoperability
  - Identity. Basic concepts, key fallacies
  - **Group work!**
- Case of Estonia
  - Big picture, identity and channels
  - X-road and infrastructure
  - governance processes

That's *my* goal, what about yours?

Architecture

On systems

- A system is, basically any combination of things that together provide more functionality than the parts alone
- Anything can be a system
  - A classroom of students
  - An aircraft
  - A field of potatoes
  - A state?
- Systems can have architecture
- Does that mean a state can have architecture?

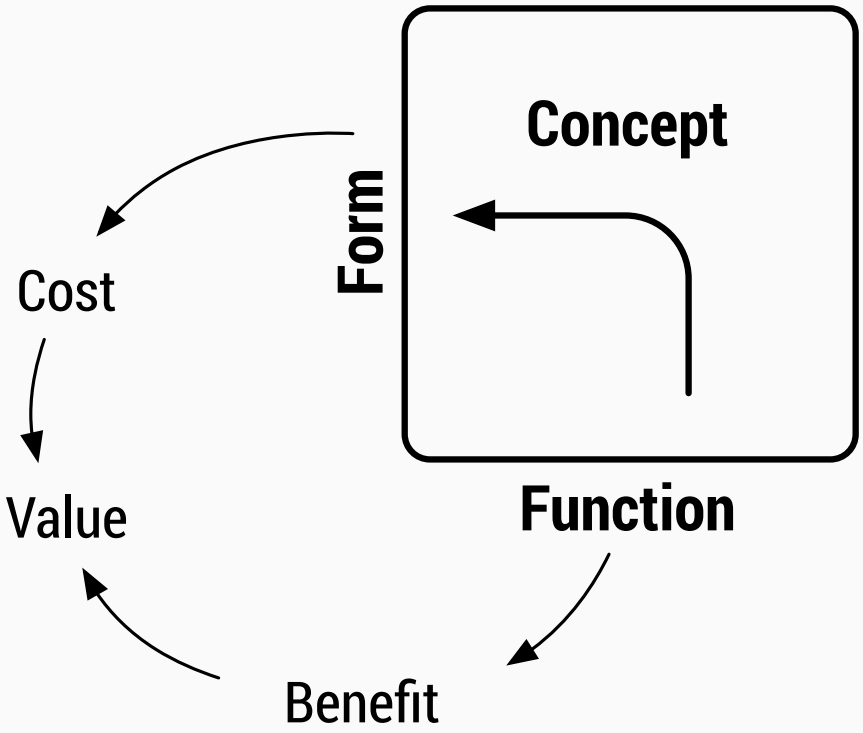
Three aspects of systems

Function is what the system *does*

Concept is the *mental model* of arriving at the form

Form is what the system *is*

Function, concept, form

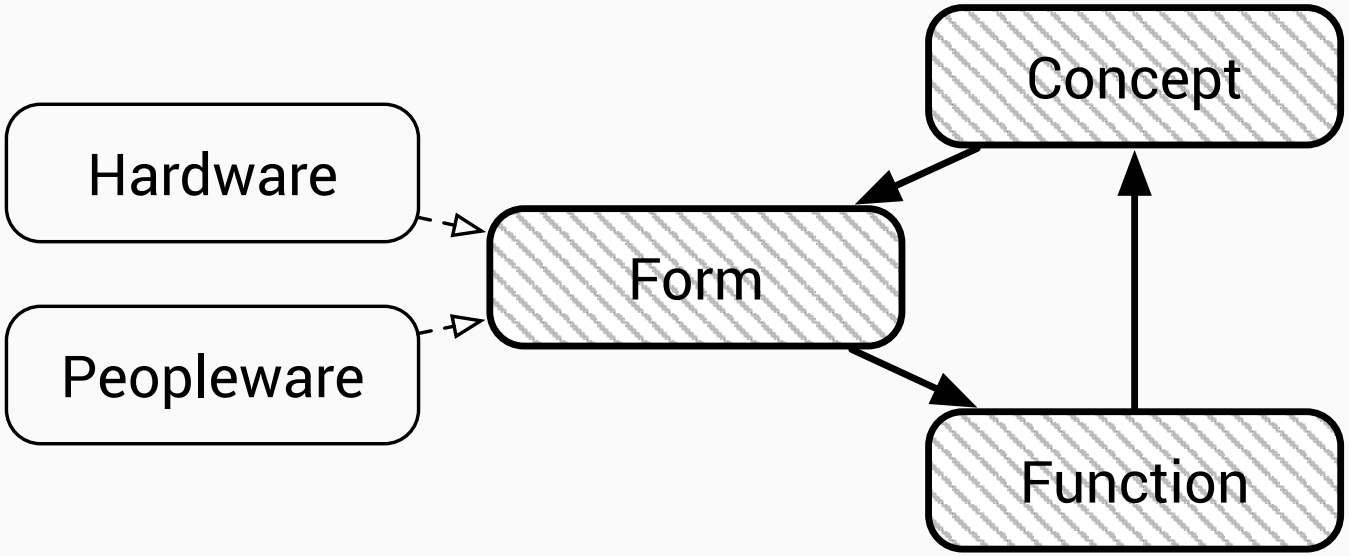


Applying the model to a state

- What is the function of a state?
- How do we think of the state and who is that “we“?
- What is the form of a state?

Key question: **what is and what is not the state?**

Traditional model of the state

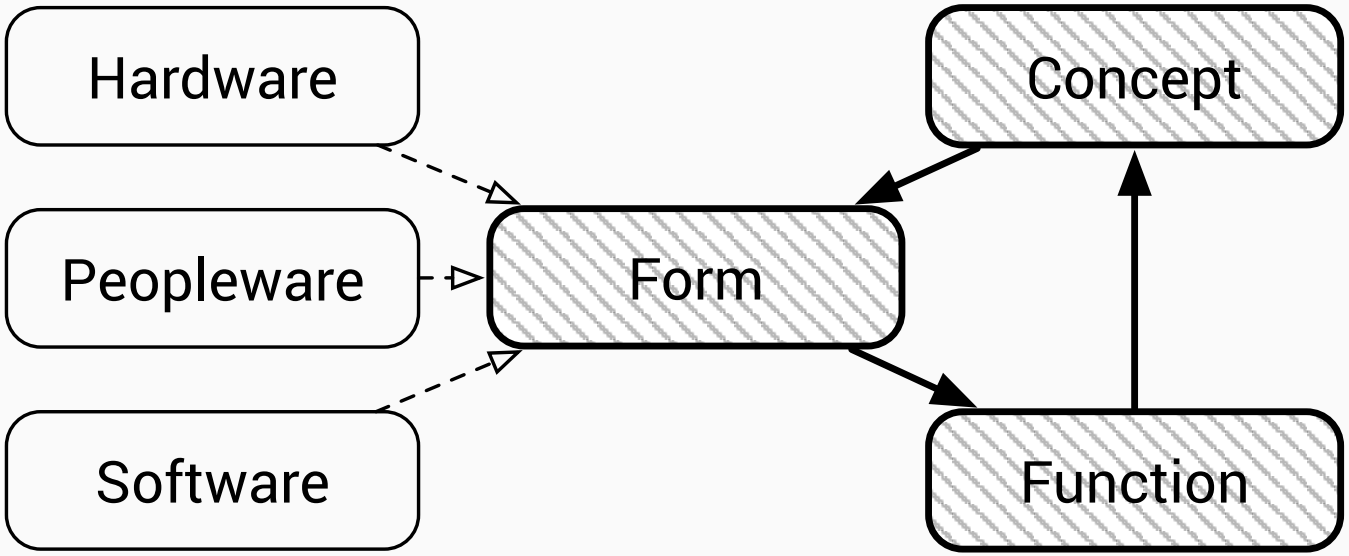


Traditional model of the state

- Based on the Westphalian nation-state model
- Key elements are
  - Infrastructure (i.e. physical territory)
  - Peopleware (i.e. government mechanism or bureaucracy)
- Emergence is static and innovation rare because people and infrastructure are static

Key point: **the model is decreasingly relevant in a global world governed by increasingly complex mechanisms**

Estonian model of the state



Estonian model of the state

- Information systems are a first-order element of the model
- The government *is* the information system
  - Information systems deliver the functions, not support bureaucracies in doing that
  - All three elements are tightly integrated
- Emergence changes frequently and fast as software is dynamic

Key point:

This is a fundamental change in how we think about the state

What is the state made of?

Ecosystems

The island

Let's imagine an island

- There is a herd of elk there
- The elk eat lichen
- The lichen survives on sunlight and humidity from the air
- The rest we do not care about at the moment
- Both the elk and lichen procreate their natural ways

Our task: **Devise a policy and action to increase the elk population**

Try to work out the solution mathematically

Go on, I'll wait...

How do we analyse the situation, then?

A policy should rely on something more than our gut feeling

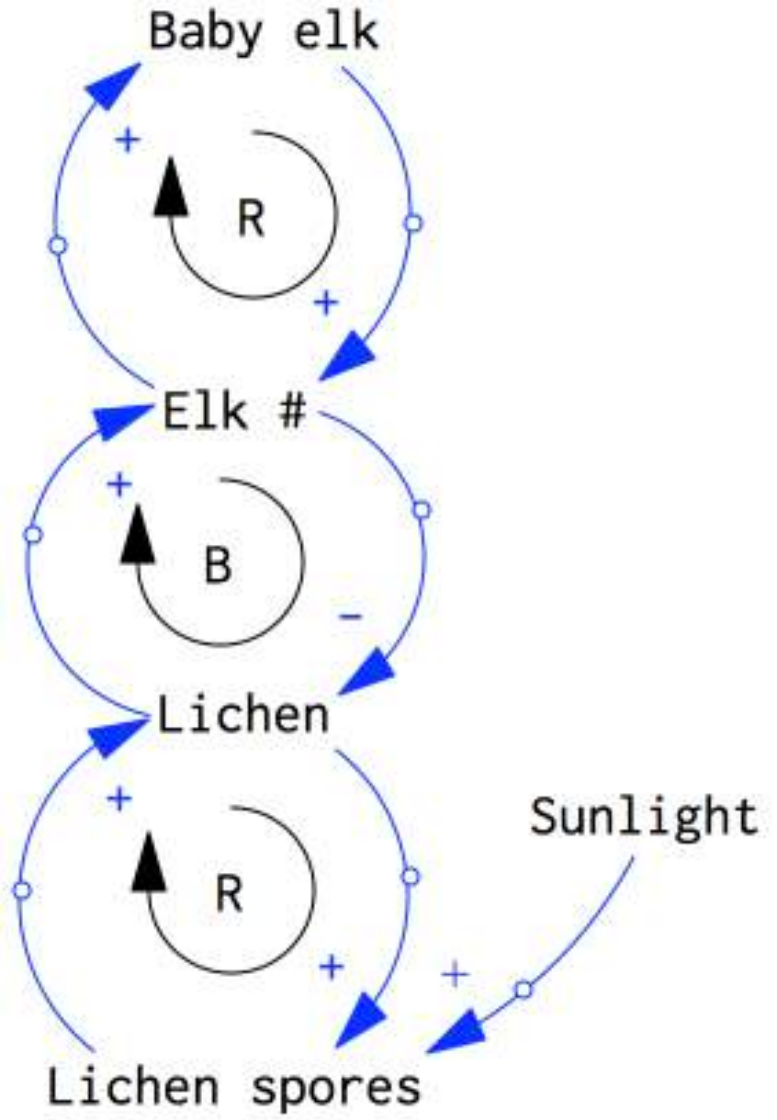
System thinking to the rescue!

- Stemming from the philosophic tradition
- Got more stringently formulated in the 1950s
- Has many offshoots: System dynamics, systems engineering/architecture, systems biology etc.
- Not to be confused with systemic approach to problems

Thinking of a problem explicitly as a system[1]

[1] A system is something that does more than its individual parts

The island ecosystem



Properties of that ecosystem

- It only takes a public good as input
  - So our elks are “free“
  - Of course we assume an infinitely large island
- It is remarkably resilient to outside disruptions
  - The ecosystems can recover automatically, if the change is small enough
  - Let’s trace what happens, if we just add to the elk population

This is what you want: sustainable progress

- No investment needed beyond initial bootstrap
- Grows rapidly until stabilisation
- Self-regulating
- Adaptive

There is no way any governance can manually keep up with technology and change

Let's add elk to the system!

That didn't work, did it?

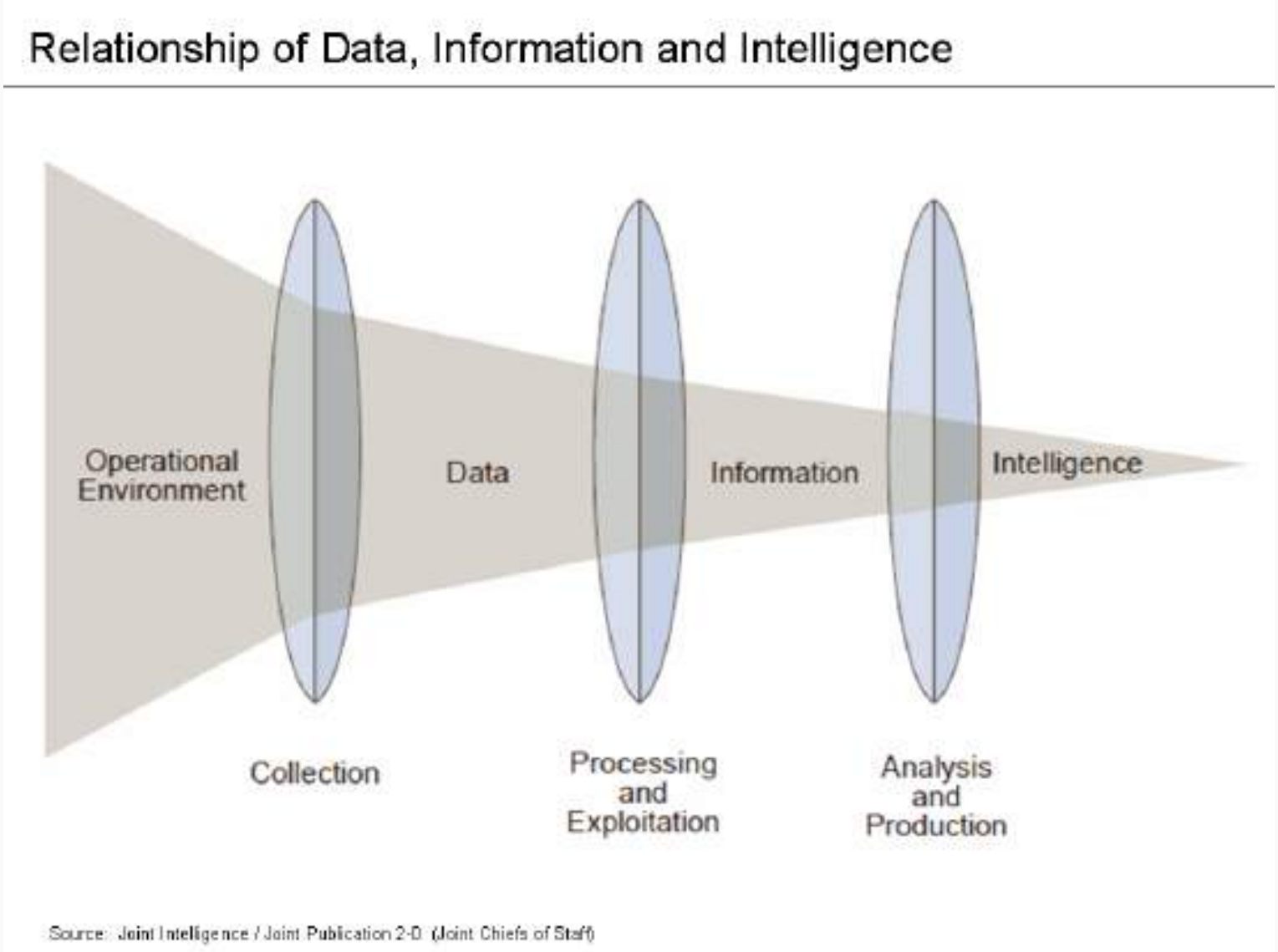
What's the solution, then?

- How do we add elk?
  - Bring in food
  - Introduce a new edible species hoping they get along with lichen
- How do we introduce elk in the first place?
  1. Understand the existing and new ecosystems (cartography)
  2. Make sure to introduce both sexes (the feedback mechanism)
  3. Make sure elk eat lichen (the support mechanism)
  4. Estimate the initial elk population (the kickstart mechanism)
  5. Ascertain the lichen only need the sun (external input)

How many e-gov ecosystems can you come up with?

Privacy

Environment, data, information, intelligence



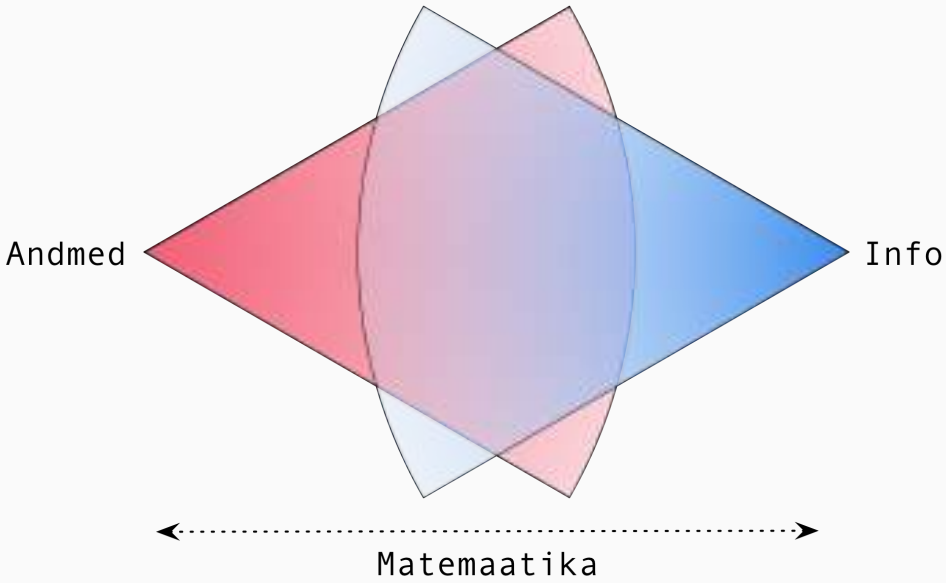
Definitions

Data is a fact. Something that is written down

Information is an answer to a question. Something you know

Intelligence is model of the situation. Something predictive

Data points and information points



What this means

Math allows free movement between data and information

- Data can be incrementally scrambled
  - Making it less and less personalised[2] and sensitive
  - The more scrambling, the less useful it is answering questions
- Information can be distilled from an array of non-sensitive information
  - The less sensitive information is, the more of it is needed
  - Information can be sensitive without being personalised

[2] Let's not go into the personal information issue

Personal data

When we are no longer dealing with personal data?

- Andres Kütt, male, id code 37508166515, weight 87 kg
- AK weighs 87 kg
- Male, lives on Kastiku street, weighs 87 kg
- 87 kg
- 666e30796e666663d46666b736d666c6373643233686663333371646e6a6c6166626676646173646173646173643233356773616
- 4e617165726620586867672c207a6e79722c207671207062717220333735303831363631352c206a72767475672038372078 74

Example: coffee

A coffee chain issuing loyalty cards and collecting a phone number

1. Who is using hashtag #COFFEE within 10 minutes after each sale with this card?
   - First sale: Probably hundreds
   - Second sale: Probably less than hundred
   - Third sale: Few, if any
2. Continue until just one username remains
3. See them post hashtag #Tartu
4. SMS: “We have a shop in Tartu, it has a sale”
5. Was there a sale in Tartu after the SMS?
6. Sell the phone number/social media combination
7. Profit

On secrecy

- There is no feasible way to fully keep even moderate secrets
- If something can get monetised, it will leak
- Nobody is interested in your data if they can get high-probability information

Therefore

- Controlled sharing of information actually enhances privacy as transparency, control and data quality are ensured
- Privacy is a process integrity problem: make sure people do with your data only what they promised to do

Summary

- Intelligence $>$ Information $>$ Data
- We protect data not because it is sensitive but because we do not know how to protect the rest
- AFAIK, generating sensitive information from non-sensitive sources *and not storing it* is not illegal[3]
- Your data is not necessary to violate your privacy
- Keep few secrets and make sure they are not valuable

[3] Terms and conditions apply, you should know this better, than I do

How does GDPR look in this context?

Identity

What we discuss here applies to everything, including people

But people are the most crucial example so we'll focus on them

Identifier

Identifier is something that uniquely identifies a person

- Examples
  - Your genetic makeup
  - For most men: name, date and location of birth
  - But not name alone or name and address
  - Your phone number
  - A username
- Identifier is always seen in a context
  - Think of how international phone numbers work
  - How many Toms were in your class and how did you solve that?
- It is only useful, if it is stable over enough time
- Uniqueness is *very* hard to guarantee

Identity

Identity is the combination of an identifier and an physical person

- Therefore, a person can have as any identities they desire
- An identity can be established without the person knowing it
  - we do this all the time
  - think cookies in a browser
- Always held together by at least one (however weak) authentication method
- Examples
  - Me and my id-code
  - Me and my skypename
  - Me and my gmail address?
- Identities can be minted based on other identities

Authentication

Authentication is the process of establishing an identity by associating a person with an identifier[4]

- Always involves an authenticating party
- Hugely complicated domain of very clever people
- Examples
  - Validation of biometry
  - Username and password
  - PIN code and possession of a card containing certificates
  - Social security number?
- All authentication is unreliable by definition
  - It is important to accurately estimate that unreliability

[4] Technically, one way is authentication and the other is identification

Identity levels

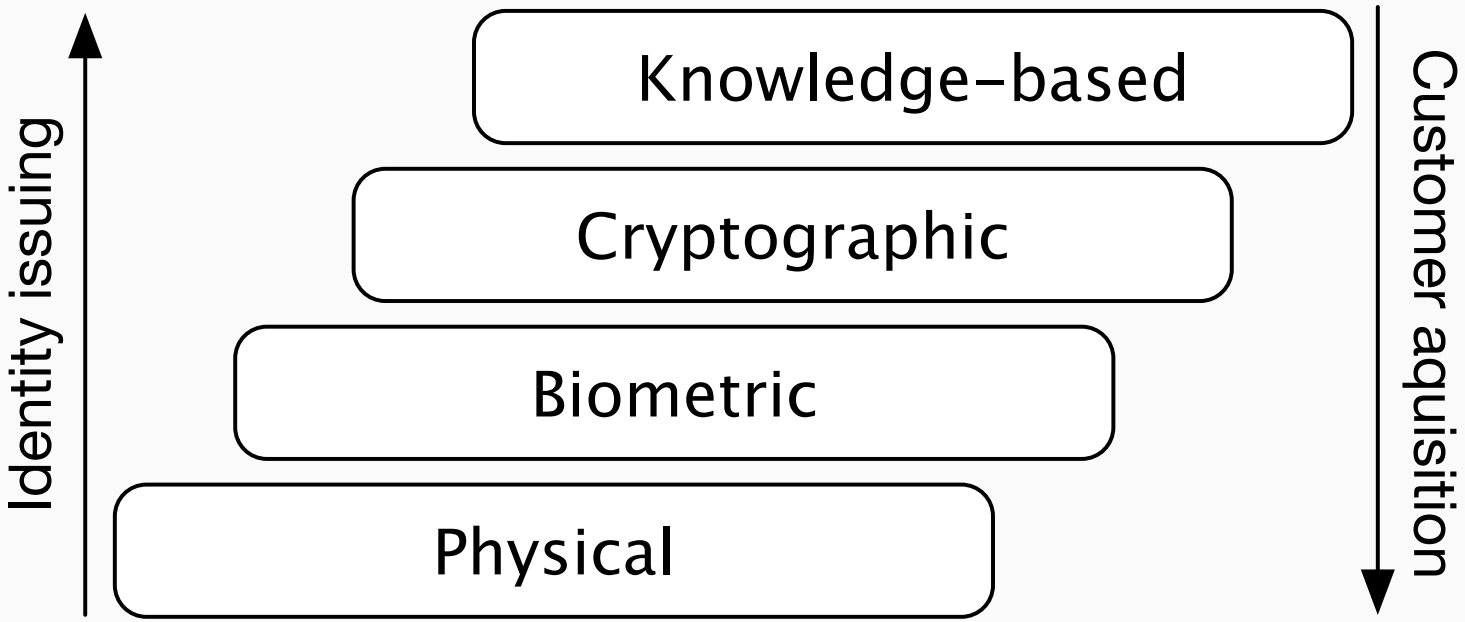
Physical Genetic makeup, your physical being

Biometric Relying on biometric authentication, possibly based on the physical identity

Cryptographic Relying on cryptographic authentication, possibly based on biometry

Knowledge-based Relying on something a person knows, possibly relying on cryptographic identity

Identity levels



Application of the correct identity level to a service is a *non-trivial* question!

Do particular legal rights and obligations apply to physical people, sets of biometry or e-mail addresses?

Moving up and down the layers

- Higher layers are cheaper and faster to issue
- Lower layers can assume very complex authentication
- Moving down
  - Use a phone number as an identifier
  - Issue certificates to the phone number
  - Call the customer into a physical location to capture biometrics
  - Establish physical identity
- Moving up
  - Every citizen gets an ID-code
  - Capture biometrics and associate with that id-code
  - Issue certificates containing the id-code
  - Issue access tokens using the certificate

Are the bottom two layers of identity necessary?

Groupwork time!

Design an ecosystem around the once only principle from the Tallinn eGov Declaration

The goal is to have an ecosystem that does not require enforcement of the principle

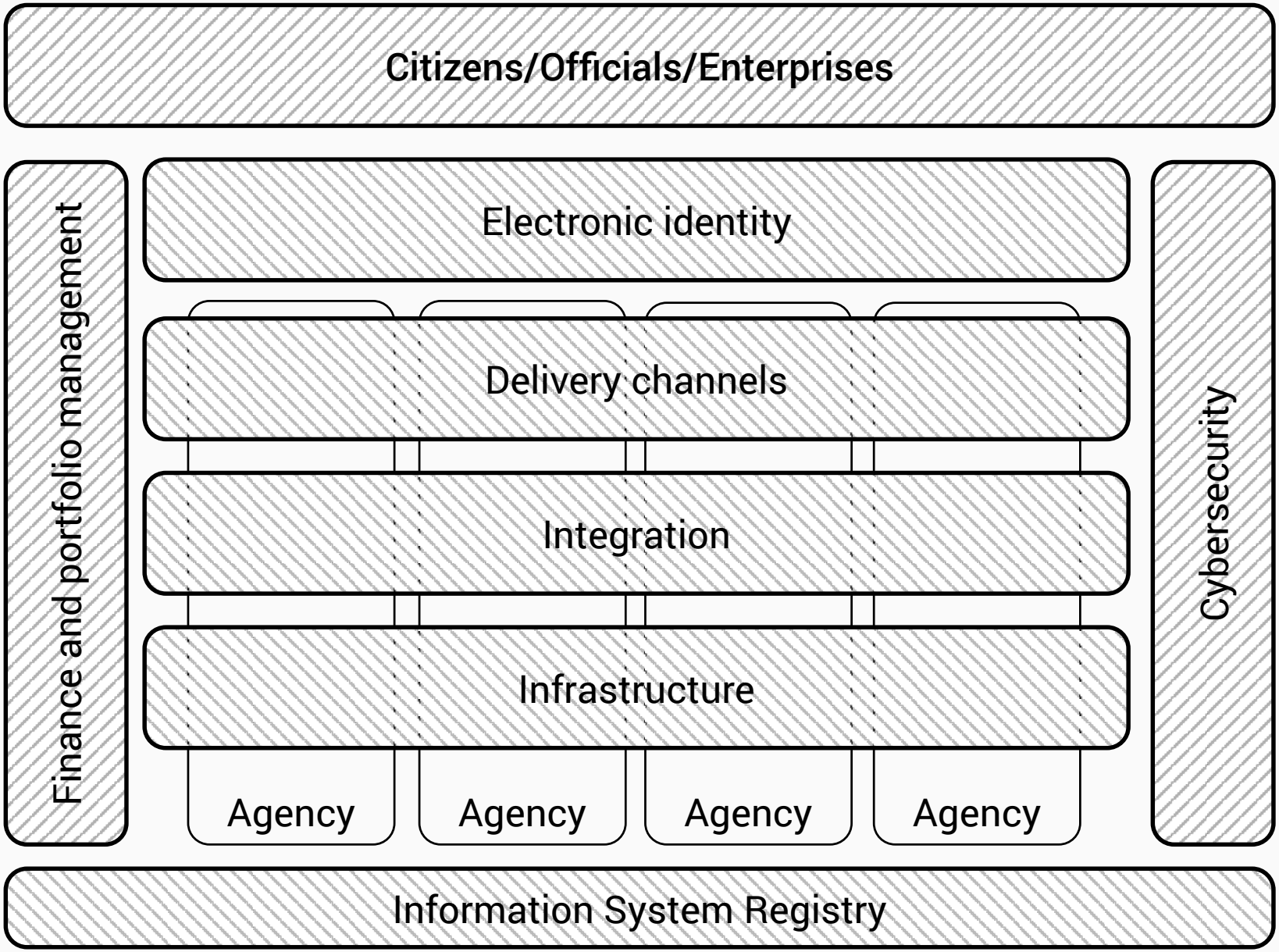
The task

You should have all the knowledge necessary

- Members of the ecosystem are
- Flashback: How do we introduce elk into the ecosystem?
  1. Understand the existing and new ecosystems (cartography)
  2. How does more “once only” lead to more “once only”? (the feedback mechanism)
  3. What is needed to keep the cycle going? (the support mechanism)
  4. How do we kickstart the cycle? (the kickstart mechanism)
  5. What external input is needed to sustain the ecosystem? (external input)
- Go through the steps even if you do not find a solid answer

How did you do?

Estonian digital architecture


Citizens/Officials/Enterprises
Finance and portfolio management
Electronic identity
Delivery channels
Integration
Infrastructure
Agency
Agency
Agency
Agency
Cybersecurity
Information System Registry

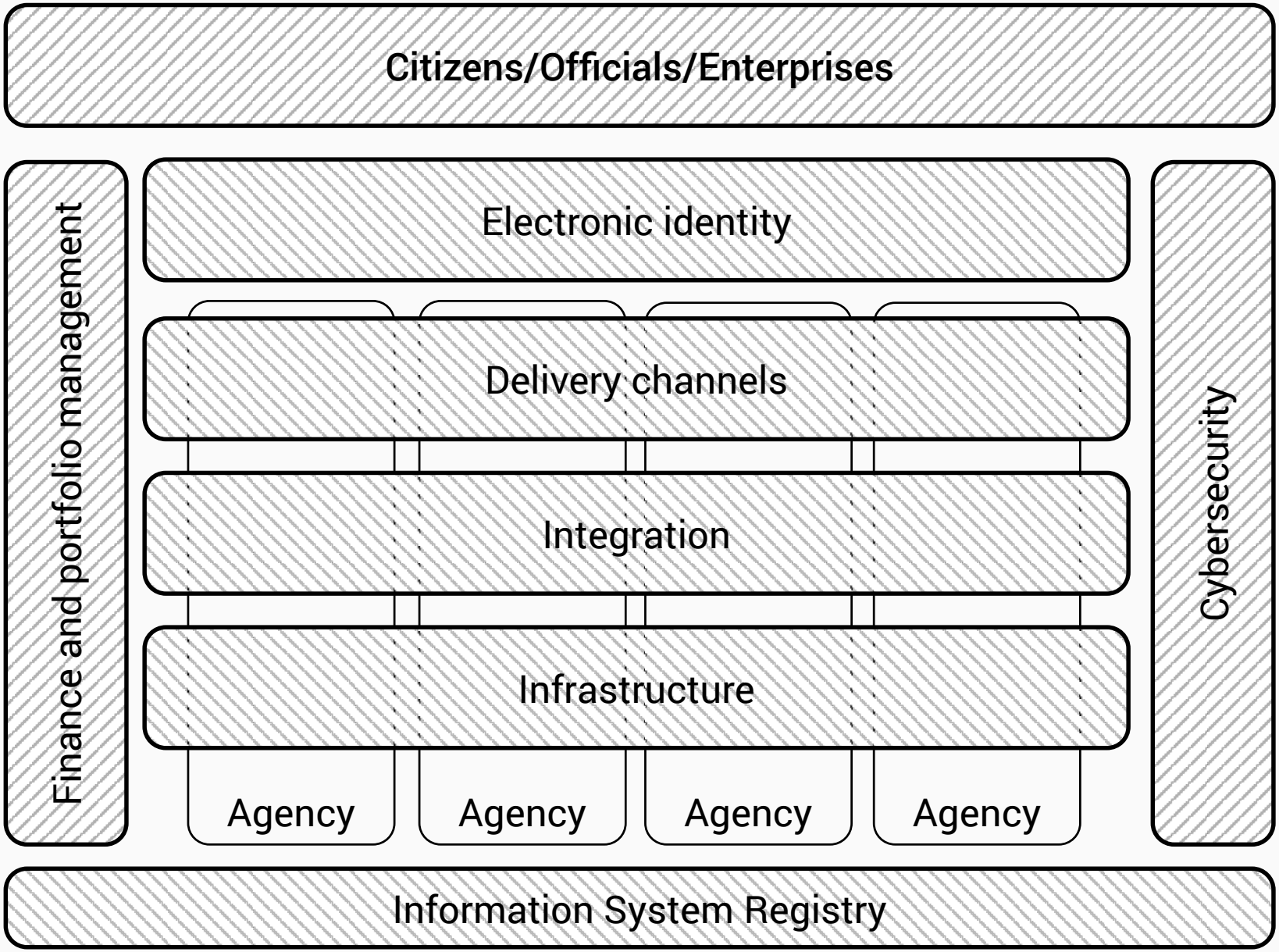
Electronic identity

- Implemented using PKI, CA service provided externally
- The certificates live on a chip (smart card or SIM)
- Several supplementary authentication methods exist, including cryptographic ones
- Digital signature legally equivalent to a physical one
- The entire system is very dependent on the id-code
- Main challenges
  - Gracefully manage the ecosystem fragmentation as new authentication methods surpass the id-card
  - eIDAS is bringing a big change in the field

Channels

- Central service portal eesti.ee with 800+ services accessible
  - Declining: it is being shut down
  - Relies on services from the next layer
  - In addition, hundreds of direct contact points with authorities
- Main challenges
  - simultaneously maintaining service ownership and central coordination
  - making people think in terms of customers
  - getting the ecosystem going
- No central UI/UX guidelines although a recommended web site template exists
- Mobile is very small but growing

What could be done about the channels ecosystem?


Citizens/Officials/Enterprises
Finance and portfolio management
Electronic identity
Delivery channels
Integration
Infrastructure
Agency
Agency
Agency
Agency
Cybersecurity
Information System Registry

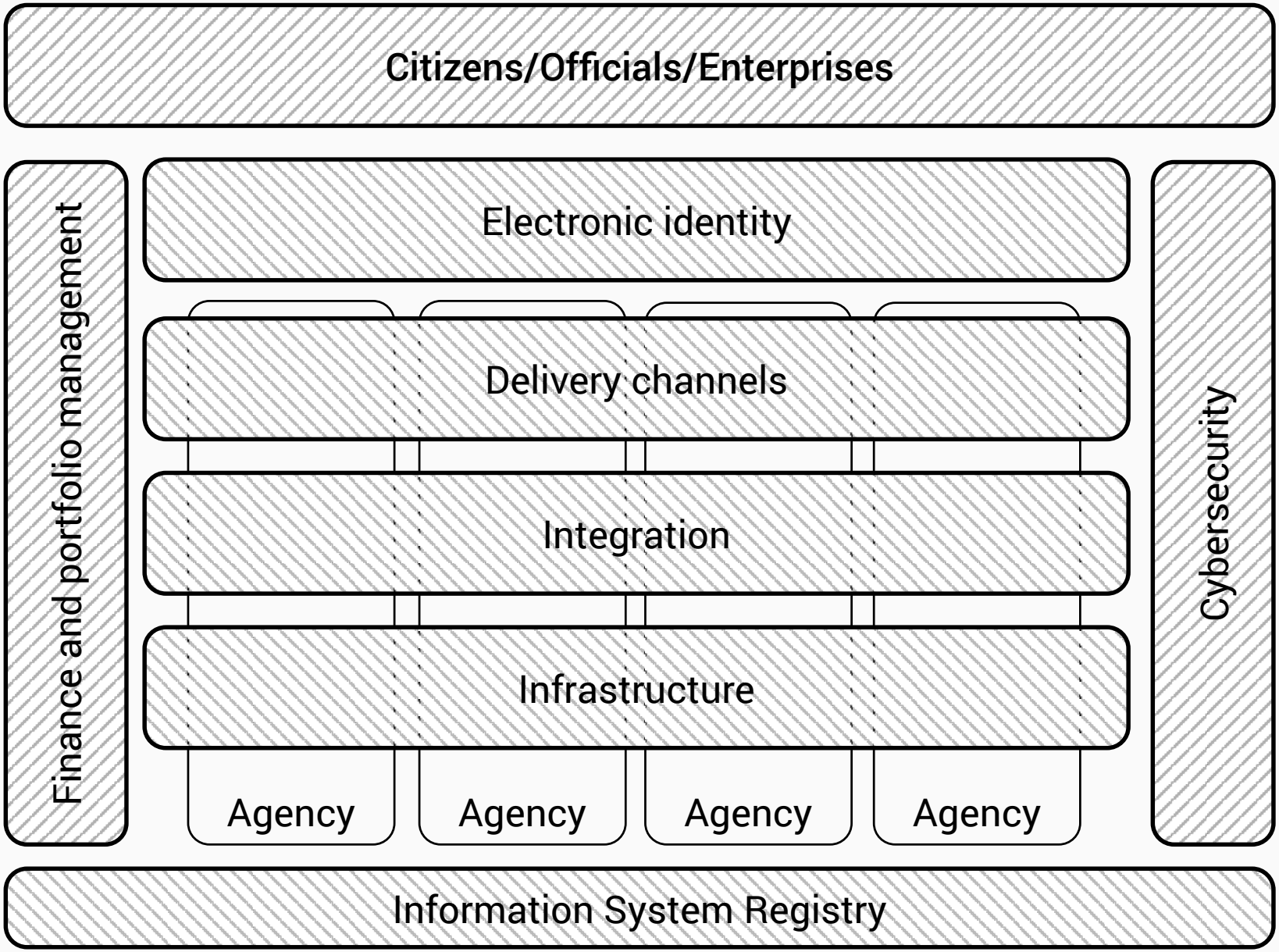
Integration

- Distributed service bus called x-road
  - all communication happens peer to peer
  - no central authority with access to traffic
  - no central development/operations bottleneck
- x-road provides standardised
  - channel crypto
  - access/identity control
  - service discovery
  - audit logging
- Massive deployment, 1000+ usable services
- Constantly developed, version 6 being rolled out
- De facto enables once-only and privacy policies
- Key challenge: keep the very large ecosystem evolving

Infrastructure

- Being expanded
  - currently mainly consolidated network access
  - government cloud in the works
  - PaaS ecosystem as a vision
- Government cloud is a combination of
  - private cloud
  - public cloud
  - data embassies
- Security and service availability major drivers: we no longer can run this country without e-services
- Scalability and cost are also becoming an issue

What are the key enablers for X-Road?


Citizens/Officials/Enterprises
Finance and portfolio management
Electronic identity
Delivery channels
Integration
Infrastructure
Agency
Agency
Agency
Agency
Cybersecurity
Information System Registry

Information system registry

A registry of all collections of private data maintained by the government

- Contains
  - Data fields with semantics
  - Technical context including X-Road service descriptors
  - Organisational context
  - Legal context
- The registry also serves as the approval tool in the Public Data Act sense
- Main challenges
  - Data quality
  - Focus: do we serve the citizen, the agencies or the approvers? In which order?

Finance and portfolio management

- Governs over centrally managed EU funding and supplementary EE IT funding
- Does not govern whatever funding can be secured via common budget processes and external sources
- Serves as both the stick and the carrot
  - You don't get funding unless you play nice with others
  - Things that need to get done, get dedicated funds
- Main IT coordination tool: it provides invaluable insight into the agencies
- Main challenges
  - Too much money
  - People getting good at writing projects
  - People not getting the point of writing projects

Cybersecurity

- Cybersecurity is an integral part of the system
- Estonia seeks to practice security by design
  - Do not design things assuming unreasonably good security
  - Cybersec people are present at all critical architecture decisions
- Cybersec is, remarkably, at the hands of civilians
- The organisation is built as an ecosystem around the "collective brain" principle
  - Everybody in the game has a rehearsed, monitored and evaluated role
  - Your current employer matters little
  - Smarter students get invited, non-performing members not

Any final thoughts, questions, comments?

Thank you!